DOCUMENT RESUME

ED 197 300 CS 005 834

AUTHOR Arnold, Richard D.; And Others
TITLE Comparison of Comprehension Questions in Three Fourth Grade Reading Texts.
PUB DATE Sep 80
NOTE 8p.; Paper presented at the Annual Meeting of the American Psychological Association (Montreal, Canada, September 1-5, 1980).

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Basal Reading; *Classification; Grade 4; Intermediate Grades; *Reading Comprehension; Reading Diagnosis; *Reading Materials; *Reading Research; Teaching Methods; *Test Reliability; *Textbooks
IDENTIFIERS *Barrett Taxonomy

ABSTRACT

A study was conducted to determine whether interjudge reliabilities on assignments of comprehension questions to the Barrett Taxonomy categories could be attained and whether the taxonomy could be used as a criterion to assess varying levels of reading comprehension questions in various reading textbooks. A total of 623 questions from three fourth grade reading texts were analyzed by three judges according to four comprehension categories: literal, inferential, evaluation, and appreciation. Means, ranges, and standard deviations were obtained, as well as total numbers of questions from each book. Several techniques were used to analyze the data. The results showed high interjudge reliability coefficients. Descriptive data varied from text to text and the analysis of variance showed Barrett's categories to be distinct, with variation among texts. Post hoc analysis of main effects revealed significant differences only in the literal category. The findings suggest that the Barrett Taxonomy is useful to categorize comprehension questions and that teachers may use it diagnostically. (FL)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

U.S. DEPARTMENT OF HEALTH,
EDUCATION & WELFARE
NATIONAL INSTITUTE OF
EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

ED197300

Comparison of Comprehension Questions in Three Fourth Grade Reading Texts

Richard D. Arnold

Associate Professor of Education

202D Education Building

Purdue University

West Lafayette, Indiana, 47907

Coauthors: George Tsatsas

Elizabeth Wilcox

Purdue University

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Richard D. Arnold

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

CS005833

Reading is typically taught through the use of a basal series, and most reading series reflect a taxonomic model of reading comprehension (Smith and Barrett, 1974, and Davis, 1944.) One well known and widely used taxonomy of cognitive processes was developed by Bloom (1956). Sanders (1966) has applied the Bloom taxonomy to reading comprehension. However, the taxonomy developed by Thomas Barrett (Clymer, 1968) is designed specifically for reading comprehension and it is more detailed, therefore, more practical for use in analyzing the thought products involved in reading and, in particular, in improving instruction in reading comprehension.

For the study reported here, the latest revision of the Barrett Taxonomy (Smith and Barrett, 1974) was used. The current version includes four categories; literal, inferential, evaluation, and appreciation.

In contrast to the voluminous empirical data reported on the Bloom taxonomy, only one article related to the Barrett Taxonomy was found and it is not empirical in nature. Therefore, the major purposes of this study are to determine if interjudge reliabilities on assignment of questions to the Barrett categories can be attained; and, if so, can the Barrett Taxonomy be used as a criterion to assess varying levels of reading comprehension questions presented among various published reading texts.

Method

Basal texts written for fourth grade pupils from three recently published series, with differently stated purposes in the teachers' manual and differing emphases were selected for this study. The American Book Company Reading Program (1977), considers comprehension as one of seven skill strands. Comprehension skills are described in this series teachers manuals in specific ways, but they are not differentiated in terms of literal, inferential and so forth. The Ginn 720 Program (1976), considers reading as comprehending the author's message. The program has seven skill strands and is described in specific ways, including reference to literal and inferential skills. Allusions are made to evaluation and appreciation, among other categories. Barrett, the developer of the taxonomy, is a senior author of this system. The Merrill Linguistic Reading Program (1975), has evolved from a system originally not concerned with reading comprehension. It now considers comprehension to be a most important goal, but does not describe comprehension in the manual with any specificity. Rather, it is described in a more global manner.

In preparation for the study three judges received three training sessions on stories not used in the study. Criteria for each category were discussed and judges practiced on training stories not used in the study. Selections were limited to fiction.

Four stories were randomly selected from each of the fourth grade texts. Questions analyzed were questions from the teachers' manuals,

from the questions in the students' text, and relevant comprehension exercises from the appropriate accompanying workbook. Some comprehension tasks were modified to conform to the question format. A total of 623 questions were categorized by each of the three judges. Means, ranges and standard deviations, as well as total numbers of questions from each book were obtained. Several statistical techniques were used; Pearson inter-judge reliability correlations, Chi Square, and ANOVA (Winer, 1971).

Findings

Interjudge reliability coefficients varied from story to story, ranging from .04 to .86. Of the 36 coefficients obtained, 32 were significant at the .05 level of significance or beyond. (Winer, 1971). These data suggest that moderate to considerable agreement was obtained among the raters. Other interjudge reliability correlations were calculated to determine whether agreement among raters was reflected by category level (e.g. literal vs. inferential) or by total number of questions rated. Obtained coefficients ranged from -.37 to .37, with all p values being non-significant. The lack of significant correlations suggest that the judges were neither influenced by the taxonomy categories nor the number of questions they rated.

The descriptive data revealed several interesting facts, a few of which are highlighted here. The average number of questions asked per story was quite different from text to text with American Book having an average of 75.6 questions per story, Ginn having an average of 50 questions per story, and Merrill having an average of 29.25 questions per story.

As indicated in Table 1, the percentage of questions per category varied considerably from text to text. For American book, nearly sixty percent of the questions were at the literal level. Inferential questions represented over one fourth of the total questions, the smallest percentage in that category for the three texts. Ginn authors distributed over sixty percent of the questions to categories other than literal, thus they distributed the largest proportion of questions to the other three categories: inferential, evaluation and appreciation. Merrill authors, though asking the fewest questions, showed the most equal distribution of questions to the four categories. Their distribution was roughly equivalent to Ginn's.

Insert Table 1 about here

From the descriptive data it is clear that the three reading texts differ mainly in the percentage of questions asked in the literal, inferential, and appreciation categories; they are very similar in the percentage of evaluation questions asked.

A test of homogeniety of variance indicated that the stories within each of the reading series were homogeneous, thus permitting the major hypotheses to be tested.

Table 2 shows the results of the main ANOVA. An F value of 20.87, significant at the .01 level, demonstrates that each of Barrett's four categories was indeed, distinct. Similarly, an F value of 7.03 ($p < .05$) shows that variation among the reading texts was present. Interaction between the series and categories was significant at the .01 level.

Insert Table 2 about here

A post hoc analysis of simple main effects was done to determine whether the differences were due to variation among all categories of the taxonomy or were limited to one or more categories. Significant differences ($p < .001$; F, 16.12; 2 df) were revealed only for the literal category, with the other categories being non-significant. This finding may reflect the predominance of literal questions for stories in all reading texts.

Discussion

The findings regarding interjudge reliability suggest several things. The moderate to high interjudge reliability coefficients obtained suggest that raters agreed rather consistently as to which category the various questions from the stories belonged. These findings suggest that the Barrett Taxonomy is a useful classification system for analyzing differing types of comprehension questions asked. However, these data also suggest that the taxonomy is not entirely clear, and that it may be possible to improve it. It is troublesome to determine whether the difficulty of classification arises from the nature of the taxonomy or the nature of the questions themselves. We suspect the problem involves both.

When interjudge reliabilities were investigated to determine the factors influencing judges ratings involving the four categories, low coefficients were obtained ($r = -.37$ to $+.37$). Also the average percent of questions per category by the three texts was found to vary considerably. Together these data suggest the possible presence of an interaction effect between categories and raters' judgements. This may be due to the large number of literal level questions contained in all the texts (29% to 77%) in comparison to the other categories.

Results of the analysis of variance suggests rather strongly that a difference exists ($p < .01$) among the questions asked in the three texts when they are classified according to the Barrett Taxonomy. It seems rather clear that the authors of the three reading texts were assigning differing amounts of question types for teachers to ask their students. It is not possible, of course, to determine whether the authors constructed the questions because of the Barrett

Taxonomy or because of some other reason. We suspect that some authors were more heavily influenced by the Taxonomy per se, than others. In particular, the Ginn fourth grade text was written by an author team of which Barrett is a senior member.

Further, in view of the philosophical orientation of the authors of the Merrill series -- that reading is primarily a "code-breaking" process -- one would expect a qualitative as well as quantitative difference in the Merrill text. In fact, differences among the three reading texts were found ($p < .05$). The raw data demonstrate rather clearly the differences among the authors as to types of questions asked. For example, the percent of literal questions asked was 59, 45 and 37. The majority of the questions asked by the authors of all three texts is in the literal category, suggesting agreement as to the importance of teaching literal comprehension, that is convergent thinking and memory. However, the degree of emphasis among authors is strikingly different. Of course, it is impossible to say which author is "right" about what percentage of questions should be devoted to which category. For example, it seems reasonable to us that children be asked literal questions somewhere between 40 to 60 percent of the time, roughly the range of the percentages represented in the study. In the final analysis, one must admit that certainly children must be able to deal with questions at the literal level before they can proceed to higher levels of thinking. How much training they need at the fourth grade level, of course, depends on many variables, a fundamental one being the maturity level of the particular students being taught. This can only be determined, it seems, by the teacher who is interacting directly with the pupils in a diagnostic manner.

The interaction present between the texts and the categories ($p < .01$) and the resulting analysis of simple main effects revealed the source of variation was clearly at the literal level. This probably reflects the predominance of literal questions over the other three categories. Had the categories been more equally represented, as was the case in the Merrill text, the test might have been more sensitive to the interaction effects. Further speculation about the interaction seems imprudent.

In summary, we believe that the study reported suggests that the Barrett Taxonomy demonstrated promise as a tool for evaluating comprehension questions of reading materials used in schools. While the taxonomy has room for improvement, it seems to have value for determining the types of comprehension questions which are contained in reading texts. Teachers can then arrange for a proper balance among these types depending upon their assessment of pupil needs. It also seems that such a classification system can reveal basic differences among published reading materials, specifically teachers manuals, which reflect differing philosophies and emphases of the authors of such reading systems. Finally, it seems that the taxonomy might be used as a basis to assess diagnostically, how children respond to different types of comprehension questions and the types of thinking represented in the question categories.

References

American Book Reading Program. Lanterns. New York, Litton, 1977.

Bloom, B. Taxonomy of educational objectives, cognitive domain. McKay, 1956.

Clymer, T. What is "reading": some current concepts. Innovation and changes in reading instruction, H. Robinson(ed), NSSE yearbook, part II: University of Chicago press, 1968, pp. 7-29.

Davis, F. B. Fundamental factors of comprehension in reading. Psychometrika, 1944, 9, 185-197.

Ginn Reading 720 Program. Tell me how the sun rose. Lexington, Xerox, 1976.

Merrill Linguistic Reading Program. Making choices. Merrill, Columbus, 1975.

Sanders, N. M. Classroom questions: what kinds? New York: Harper, 1966.

Smith, R. J. and Barrett, T. C. Teaching reading in the middle grades. Reading, Mass: Addison Wesley, 1974.

Weiner, B. Statistical procedures in experimental design. New York: McGraw-Hill, 1971.

Table 1

Average Percentages of Question Types Presented in Three Fourth Grade Reading Texts

Publisher	Literal	Inferential	Evaluation	Appreciation
American Book	58.5	27.75	7.25	6.5
Ginn	36.5	43.25	8.0	12.25
Merrill	45.75	41.25	7.75	5.25

Table 2

Analysis of Variance

Source	SS	df	MS	F	p
Between	1840.23	11			
A (series)	1121.79	2	560.90	7.03	.05
stories within series (residual)	718.44	9	79.83		
Within	7832.75	36			
B (Barrett's categories)	4409.73	3	1469.91	20.87	.01
AB	1521.71	6	253.62	3.60	.01
BX stories within series (residual)	1901.31	27	70.42		